

# Content Moderation

## Communications & Media

Rules, scale and the two kinds of error you must trade off.

### Over-removal

- Legitimate speech taken down
- Satire and reporting caught by filters
- Minority languages served worst
- Appeals slow or unavailable
- Chilling effect on borderline topics

### Under-removal

- Real harm left up and amplified
- Targets bear the cost of reporting
- Coordinated abuse outpaces review
- Advertisers and regulators react
- Trust in the platform erodes

#### THE DIFFERENCE THAT MATTERS

Any threshold produces both errors. The policy question is which error to accept and for whom.

Educational reference. Legal duties on platforms differ sharply by jurisdiction and change often. Not legal advice.